



Statistical properties of nearest neighbor regression function estimate for strong mixing processes.

Patrick Rakotomarolahy

► To cite this version:

Patrick Rakotomarolahy. Statistical properties of nearest neighbor regression function estimate for strong mixing processes.. 2012. hal-00725841

HAL Id: hal-00725841

<https://hal.science/hal-00725841>

Submitted on 27 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical properties of nearest neighbor regression function estimate for strong mixing processes.

Patrick Rakotomarahy*

Abstract

We examine statistical properties of nearest neighbor regression technique beyond the standard i.i.d hypothesis. We analyze the second order properties and establish the asymptotic normality of the nearest neighbor regression function estimate for strongly mixing processes. We achieve some rates for the second order properties for this nonparametric method for α -mixing processes. We make an application of the theoretical results on the modelling of economic indicators.

Keywords: nearest neighbor, α -mixing processes, second order properties, asymptotic normality, economic indicators.

JEL: C22 - C53 - E32.

1 Introduction

Properties of the k nearest neighbor (k -NN) regression estimate depend on the dependence structure of the process. Under the *i.i.d* assumption, many results have been obtained about the properties of the k -NN regression function estimate. Not to be exhaustive, we can mention the work of Royall (1966) on k -NN regression estimate with uniform weighting function. He studied the MSE, the MISE and the asymptotic normality of such k -NN regression. Later, Mack (1981) extended this result for non-uniformly weighted k -NN regression estimate. He investigated the bias and the asymptotic normality of the k -NN regression estimate with a more general weighting function, again under the *i.i.d* assumption. Other types of convergence such as L^2 consistent and

*Centre d'Economie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne, 106 boulevard de l'Hopital 75647 Paris Cedex 13, France, e-mail: rakotopapa@yahoo.fr.

uniform convergence are discussed in Stone (1977) and Devroye (1982), respectively. Moreover, Stute(1984) has weakened the assumption on finiteness of the third moment of the underlying process in Mack (1981) and has obtained the asymptotic normality of the univariate k -NN regression estimate.

An interesting extension would be on relaxing the *i.i.d* assumption, may we still establish similar statistical properties of the k -NN regression estimate even the processes are dependent. That would ensure the use of k -NN estimate to time series processes.

In contrast to independent framework, fewer results on k -NN regression estimate have been obtained for dependent processes. They are mainly obtained in univariate setting. Among them, we can mention the work of Collomb (1984) where he has provided the piecewise convergence for univariate dependent processes. This work has been extended by Yakowitz (1987) on the quadratic mean square error of such k -NN regression estimate. We will enlarge these results on dependent processes for the k -NN regression estimate. Clearly, we will analyze the bias, the variance and the quadratic mean square error of the k -NN estimate and second. We will also establish the asymptotic normality of such regression estimate.

The paper is organized as follows. Section 2 introduces the k -NN nonparametric regression function estimate. Section 3 studies properties of this nonparametric regression function estimate for dependent processes. The last section provides an application of the theoretical result on real data.

2 Nearest neighbor regression estimate

Let provide a brief presentation of a k -NN regression estimate of the unknown conditional mean. Consider a process (Y, \underline{X}) with valued in $\mathbb{R} \times \mathbb{R}^d$. We assume that the process has a joint density function f . For $\underline{x} \in \mathbb{R}^d$, we are interested in estimation of the conditional mean $m(\underline{x}) = E[Y \mid \underline{X} = \underline{x}]$.

Given a sample $(Y_1, \underline{X}_1), \dots, (Y_n, \underline{X}_n)$, with size n , for the process (Y, \underline{X}) , the k -NN regression

estimate of the conditional mean $m(\underline{x})$ can be rewritten as follows:

$$m_n(\underline{x}) = \sum_{i=1}^n w(\underline{x} - \underline{X}_i) Y_i \quad (2.1)$$

where w is a weighting function associated to neighbors. A general expression of the weighting function w is:

$$w(\underline{x} - \underline{X}_i) = \frac{\frac{1}{nR_n^d} K\left(\frac{\underline{x} - \underline{X}_i}{R_n}\right)}{\frac{1}{nR_n^d} \sum_{i=1}^n K\left(\frac{\underline{x} - \underline{X}_i}{R_n}\right)} \quad (2.2)$$

where R_n will be defined as a distance, according to the Euclidean norm in \mathbb{R}^d , from \underline{x} to its k -th neighbors, and $K(u)$ is a bounded, non negative function satisfying

$$\int K(u) du = 1 \quad \text{and} \quad K(u) = 0 \quad \text{for} \quad |u| \geq 1. \quad (2.3)$$

We wish to extend results for independent processes to dependent processes. Let us remind the notion of dependence. The concept of dependence may occur in different ways according to the structure of the process to be studied and could be ranged from the weakest to the strongest. Here, we consider some mixing conditions which measure the dependence structure on the process. They are the strong, the uniform and the regular mixings. Consider a process $\{X_i, i \in \mathbb{N}\}$ which is assumed to be stationary and denote by $\mathcal{F}_r^s = \sigma(X_\ell, r \leq \ell \leq s)$ the σ -algebra generated by $\{X_\ell, r \leq \ell \leq s\}$. We define the following mixing coefficients:

$$\alpha(n) = \sup_{A \in \mathcal{F}_1^t, B \in \mathcal{F}_{n+t}^\infty} |P(B \cap A) - P(A)P(B)|. \quad (2.4)$$

$$\beta(n) = \sup_{B \in \mathcal{F}_{n+t}^\infty} |P(B) - P(B | \mathcal{F}_1^t)|. \quad (2.5)$$

$$\phi(n) = \sup_{A \in \mathcal{F}_1^t, B \in \mathcal{F}_{n+t}^\infty, P(A) \neq 0} |P(B | A) - P(B)|. \quad (2.6)$$

Then the process $\{X_i, i \in \mathbb{N}\}$ is said to be α -mixing (or strong mixing) if $\alpha(n) \rightarrow 0$ as $n \rightarrow \infty$, β -mixing if $\beta(n) \rightarrow 0$ as $n \rightarrow \infty$ and ϕ -mixing (or uniform mixing) if $\phi(n) \rightarrow 0$ as $n \rightarrow \infty$. Moreover, such process is said to be geometrically strongly mixing if there exist $c_0 > 0$ and $\rho \in [0, 1[$ such that $\alpha(n) \leq c_0 \rho^n$. For more development on the concepts of dependent processes, Rosenblatt (1956), Ibragimov (1962), Peligrad (1986), Bradley (1986), Doukhan (1994) and Bosq (1998). Example of dependent processes are the linear and the nonlinear parametric models, since processes like the ARMA models, related GARCH processes and Markov switching processes are known to be mixing, Davydov(1973), Guégan (1983), Mokkadem(1990), and Carrasco and Chen

(2002), among others.

We reconstruct the process (Y, \underline{X}) using embedding principle which allows taking into account some characteristics of the series which are not always observed on the trajectory in \mathbb{R} . We consider a real time series $(X_n)_n$, observed until time n . We transform the original process by embedding it in a space of dimension d and therefore we have a new process $\underline{X} = (\underline{X}_n)_n$ with valued in \mathbb{R}^d where $\underline{X}_i = (X_{i-d+1}, \dots, X_i)$, $i = d, \dots, n$, and if not mentioned $Y_i = X_{i+1}$. Then, we have (Y_i, \underline{X}_i) with $i = d, \dots, n$, realizations of the process (Y, \underline{X}) . For a given distance measure and $\underline{x} \in \mathbb{R}^d$, we compute the distance between \underline{x} and \underline{X}_i for $i = d, \dots, n$. So, we can define a neighborhood around \underline{x} , $N(\underline{x}) = \{i \mid (Y_{(i)}, \underline{X}_{(i)}), i = 1, \dots, k\}$ whose $\underline{X}_{(i)}$ represents the i -th nearest neighbor of \underline{x} . We assume that the process $(X_n)_n$ is stationary. Moreover, we assume that the random variable $Y_n \mid (\underline{X}_n = \underline{x})$ has a conditional density $f(y \mid \underline{x})$, and the invariant measure associated to the embedded time series $(\underline{X}_n)_n$ is h .

Our first result concerns the bias and the quadratic mean square error of the k -NN regression estimate of $m(\cdot)$. For strongly mixing processes with suitable conditions (on the density functions, on the weighting function and on the regression function), the bias and the quadratic mean square error of the k -NN regression estimate are of order $O(n^{-\beta})$ and of order $O(n^{-Q})$, respectively with $\beta = \frac{(1-Q)p}{d}$, $Q = \frac{2p}{2p+d}$ and p the degree of smoothness. The next result is on the asymptotic normality of the same k -NN regression estimate. Under suitable conditions, the centered k -NN regression estimate scaled by the corresponding asymptotic variance has asymptotic standard normal distribution with speed rate of convergence $\sqrt{n^Q}$.

2.1 Second order properties

Second order properties analyzed in this section concern the bias, the variance and the quadratic mean square error of the k -NN regression estimate $m_n(\cdot)$ in relation (2.1). Before all, we introduce some assumptions. The first assumption characterizes the dependence on the process.

Assumption 2.1. $(X_n)_n$ is α -mixing process. Moreover, for a certain $\delta > 0$, $|X_n|^{2+\delta}$ is uniformly integrable, $\sum_n n^{2/\delta} \alpha(n) < \infty$ and $\inf_k \text{Var}(X_k) > 0$

The next assumption concerns regularity condition on the joint, on the marginal and on the

conditional density functions of the process.

Assumption 2.2. $h(\underline{x}), f(y \mid \underline{x})$ and $f(y, \underline{x})$ are p continuously differentiable and $f(y \mid \underline{x})$ is bounded.

The last assumption, for deriving second order properties of k -NN regression estimate, states conditions on the number of neighbors k and on the weighting function.

Assumption 2.3. The sequence $k(n) < n$ is such that $\sum_{i=1}^{k(n)} w_i = 1$ where w_i a weighting function satisfying $0 < w_i < 1$ when $i \leq k(n)$ and $w_i = 0$ otherwise.

The regularity condition on the marginal and conditional distribution of the time series in assumption 2.2 can be overcome when the function $m(\underline{x})$ satisfies Lipschitz's condition with order $1 \leq \delta < 2$.

We now present the main results on the k -NN regression estimate for dependent processes. The result established in theorem 2.1 provides second order properties of the k -NN regression estimate.

Theorem 2.1. If assumptions 2.1-2.3 hold, then :

(i) The k -NN regression estimate $m_n(\underline{x})$ in relation 2.1 is asymptotically unbiased

$$E[m_n(\underline{x})] = m(\underline{x}) + O(n^{-\beta}). \quad (2.7)$$

(ii) The variance of the k -NN regression estimate $m_n(\underline{x})$ is given by:

$$\text{Var}[(m_n(\underline{x}))] = \gamma^2 \left(\frac{v(\underline{x})}{k(n)} \right) \quad (2.8)$$

(iii) The quadratic mean square error of the k -NN regression estimate $m_n(\underline{x})$ satisfies

$$E[(m_n(\underline{x}) - m(\underline{x}))^2] = O(n^{-Q}), \quad (2.9)$$

where $0 \leq Q < 1$, $Q = \frac{2p}{2p+d}$, $\beta = \frac{(1-Q)p}{d}$, $v(\underline{x}) = \text{Var}(Y_n / X_n = \underline{x})$ and γ is a positive constant which is equal to 1 when we use uniform weights.

The basic idea of the proof is as follows. In general, the number of neighbors k depends on the sample size n and should be written $k(n)$. But we will just write k instead of $k(n)$. This integer

number k is the key unknown parameter of the k -NN regression estimate. Large-sample properties should account appropriate value for the unknown number k . The choice of the number of neighbors $k = \lfloor n^Q \rfloor$, with $0 \leq Q < 1$ and $\lfloor \cdot \rfloor$ the integer part symbol, is frequent for matching the usual conditions ($k = o(n)$, $\log(n) = o(k)$ and $k \rightarrow \infty$ as $n \rightarrow \infty$) for large-sample properties such as for derivation of convergence rate or asymptotic distribution of the k -NN regression estimate. Then, calibration of the number k is equivalent to the characterization Q . We calibrate the parameter k based on the bias-variance dilemma. Therefore, for dependent processes, we establish second order properties of the k -NN regression method.

Proof of theorem 2.1. We start from providing the proof of (i) which is the bias of the k -NN regression estimate $m_n(\underline{x})$. We denote $B(\underline{x}, r_0) = \{z \in \mathbb{R}^d, \|\underline{x} - z\| \leq r_0\}$ the ball centered at \underline{x} with radius $r_0 > 0$. We characterize the radius r insuring that $k(n)$ observations fall in the ball $B(\underline{x}, r)$; indeed, since the function $h(\cdot)$ is p -continuously differentiable, for a given i the probability q_i of an observation \underline{x}_i to fall in $B(\underline{x}, r)$ is:

$$q_i = P(\underline{x}_i \in B(\underline{x}, r)) \quad (2.10)$$

$$= \int_{B(\underline{x}, r)} h(\underline{x}_i) d\underline{x}_i = h(\underline{x}) \cdot \int_{B(\underline{x}, r)} d\underline{x}_i + \int_{B(\underline{x}, r)} (h(\underline{x}_i) - h(\underline{x})) d\underline{x}_i \quad (2.11)$$

$$= h(\underline{x}) c r^d + o(r^d), \quad (2.12)$$

where c is the volume of the unit ball and $d\underline{x} = dx_1 dx_2 \cdots dx_d$. Thus, $q_i - q_j = o(r^d)$ for all $i \neq j$. We consider now the k -NN vectors $\underline{x}_{(k)}$ and we denote q the probability that they are in the ball $B(\underline{x}, r)$, that is $q = P(\underline{x}_{(k)} \in B(\underline{x}, r))$, then :

$$q_i = q + o(r^d). \quad (2.13)$$

Being given $N(r, n)$, the number of observations falling in the ball $B(\underline{x}, r)$, for a given $r > 0$, we characterize r such that $k(n)$ observations fall in $B(\underline{x}, r)$. We proceed as follows. We denote \mathcal{S}_i^n

all non ordered combinations of the i -uple indices from $(n - d)$ indices, then:

$$\begin{aligned}
E[N(r, n)] &= \sum_{i=0}^{n-d} i P(N(r, n) = i) = \sum_{i=0}^{n-d} i \sum_{(j_1, \dots, j_i) \in \mathcal{S}_i^n} \prod_{j=j_1}^{j_i} q_j \prod_{\substack{\ell=1 \\ \ell \notin \{j_1, \dots, j_i\}}}^{n-d} (1 - q_\ell) \\
&\geq \sum_{i=0}^{n-d} i \sum_{(j_1, \dots, j_i) \in \mathcal{S}_i^n} \underline{q}^i (1 - \bar{q})^{n-d-i} = \sum_{i=0}^{n-d} i \binom{n-d}{i} \underline{q}^i (1 - \bar{q})^{n-d-i} \quad (2.14)
\end{aligned}$$

$$= \underline{q}(n-d)(1 + \underline{q} - \bar{q})^{n-d},$$

where \underline{q} and \bar{q} are respectively the smallest and largest probabilities q_i $i = 1, \dots, n-d$. Thus, we obtain a lower bound for $E[N(r, n)]$. If $E[N(r, n)] = k(n)$, using (2.12) - (2.14), we obtain:

$$r \leq \left(\frac{k(n)}{(n-d)} \right)^{\frac{1}{d}} D(\underline{x}), \quad (2.15)$$

with $D(\underline{x}) = \left(\frac{1}{h(\underline{x})c} \right)^{\frac{1}{d}}$.

Now, using the relationship (2.1), we get:

$$E[m_n(\underline{x})] = \sum_{i \in N(\underline{x})} E[w(\underline{x} - \underline{X}_{(i)}) Y_i], \quad (2.16)$$

where $Y_i = X_{(i)+1}$. We can remark that $E[w(\underline{x} - \underline{X}_{(i)}) Y_i] = \int_{\mathbb{R}^d} \int_{\mathbb{R}} w(\underline{x} - \underline{x}_i) y_i f(y_i, \underline{x}_i) d\underline{x}_i dy_i$. Since $f(y_i, \underline{x}_i) = f(y_i | \underline{x}_i) h(\underline{x}_i)$, then we obtain $E[w(\underline{x} - \underline{X}_{(i)}) Y_i] = \int_{\mathbb{R}^d} \int_{\mathbb{R}} w(\underline{x} - \underline{x}_i) y_i f(y_i | \underline{x}_i) h(\underline{x}_i) d\underline{x}_i dy_i$. Thus, as soon as the weighting function $w(\cdot)$ is vanishing outside the ball $B(\underline{x}, r)$:

$$E[w(\underline{x} - \underline{X}_{(i)}) Y_i] = \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) \left(\int_{\mathbb{R}} y_i f(y_i | \underline{x}_i) dy_i \right) h(\underline{x}_i) d\underline{x}_i \quad (2.17)$$

$$= \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) m(\underline{x}_i) h(\underline{x}_i) d\underline{x}_i. \quad (2.18)$$

To compute the bias we need to evaluate: $E[m_n(\underline{x})] - m(\underline{x})$. We begin to evaluate :

$$\sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) m(\underline{x}_i) h(\underline{x}_i) d\underline{x}_i = m(\underline{x}) E\left[\sum_{i \in N(\underline{x})} w(\underline{x} - \underline{X}_{(i)}) \right] = m(\underline{x}). \quad (2.19)$$

Then,

$$E[m_n(\underline{x})] - m(\underline{x}) = \sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) (m(\underline{x}_i) - m(\underline{x})) h(\underline{x}_i) d\underline{x}_i. \quad (2.20)$$

The equation (2.20) holds because $\sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) h(\underline{x}_i) d\underline{x}_i = 1$, (Assumption 2.3). Then,

$$|E[m_n(\underline{x})] - m(\underline{x})| \leq \sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) a \|\underline{x}_i - \underline{x}\|^p h(\underline{x}_i) d\underline{x}_i. \quad (2.21)$$

We get this last expression since the constant a is known and $m(\cdot)$ is p -continuously differentiable. The inequality (2.21) implies that:

$$|E[m_n(\underline{x})] - m(\underline{x})| \leq ar^p E\left[\sum_{i \in N(\underline{x})} w(\underline{x} - \underline{X}_{(i)})\right]. \quad (2.22)$$

The relationship in (2.22) holds because $\|\underline{x}_i - \underline{x}\|^p < r^p$, as soon as $\underline{x}_i \in B(\underline{x}, r)$. Now, both cases be considered:

1. When r is very small, than the bias is negligible and $E[m_n(\underline{x})] = m(\underline{x})$.
2. If the bias is not negligible, using (2.15) and (2.22), we get:

$$|E[m_n(\underline{x})] - m(\underline{x})| \leq a \left(\frac{k(n)}{(n-d)} \right)^{\frac{p}{d}} D(\underline{x})^p. \quad (2.23)$$

If we choose $k(n)$ as in integer part of n^Q , and knowing that $\frac{k}{n-d} \sim \frac{k}{n}$, then $|E[m_n(\underline{x})] - m(\underline{x})| = O(n^{-\beta})$ with $\beta = \frac{(1-Q)p}{d}$.

The proof of theorem 2.1 point (i) is complete. Next we prove point (ii) of theorem 2.1 which concerns the variance of the k -NN regression estimate $m_n(\underline{x})$.

The variance of $m_n(\underline{x})$ can be written as follows:

$$\begin{aligned} Var(m_n(\underline{x})) &= \sum_{i=1}^{k(n)} Var(w(\underline{x} - \underline{X}_{(i)})Y_i) + \sum_{i=1}^{k(n)} \sum_{j \neq i}^{k(n)} cov(w(\underline{x} - \underline{X}_{(i)})Y_i, w(\underline{x} - \underline{X}_{(j)})Y_j) \\ &= \frac{\gamma^2}{k(n)^2} \sum_{i=1}^{k(n)} Var(Y_i) + \frac{\gamma^2}{k(n)^2} \sum_{i=1}^{k(n)} \sum_{j \neq i}^{k(n)} cov(Y_i, Y_j) \end{aligned} \quad (2.24)$$

Using first the inequality on the covariance of α -mixing processes in Lin and Lu (1996) and next

the stationary condition, relation (2.24) becomes:

$$\begin{aligned}
\text{Var}(m_n(\underline{x})) &\leq \frac{\gamma^2}{k(n)^2} \sum_{i=1}^{k(n)} \text{Var}(Y_i) + \frac{\gamma^2}{k(n)^2} \sum_{i=1}^{k(n)} \sum_{j \neq i}^{k(n)} \alpha(|(i) - (j)|) (E[Y_i^2])^{1/2} (E[Y_j^2])^{1/2} \\
&\leq \frac{\gamma^2}{k(n)^2} \left(\sum_{i=1}^{k(n)} E[Y_i^2] + \sum_{i=1}^{k(n)} \sum_{j \neq i}^{k(n)} \alpha(|(i) - (j)|) (E[Y_i^2])^{1/2} (E[Y_j^2])^{1/2} \right) \\
&\leq \gamma^2 \left(\frac{1}{k(n)} + \frac{1}{k(n)^2} \sum_{i=1}^{k(n)} \sum_{j \neq i}^{k(n)} \alpha(|(i) - (j)|) E[Y_i^2] \right)
\end{aligned} \tag{2.25}$$

From assumption 2.1, we get $\sum_{i=1}^{k(n)} \sum_{j \neq i}^{k(n)} \alpha(|(i) - (j)|) < \infty$ as n tends to infinity. When working on centered time series, the following equality always holds: $E[Y_i^2] = \text{Var}(Y_i)$. Moreover, we know that $\text{Var}(Y_i) = \text{Var}(m(\underline{X}_{(i)})) + E(v(\underline{X}_{(i)}))$. Since $\underline{X}_{(i)} \in B(\underline{x}, r)$, then from Taylor expansion around \underline{x} of $m(\cdot)$ and $v(\cdot)$, we get $\text{Var}(Y_i) = v(\underline{x}) + O(r)$. Relation (2.25) yields a characterization of the variance as follows: $\text{Var}(m_n(\underline{x})) = \gamma^2 \frac{1}{k(n)} (v(\underline{x}) + O(r))$. We have just specified the variance of the estimate m_n . We now prove (iii) of theorem 2.1.

We derive the proof of (iii) from (i) and (ii) of theorem 2.1. We start by writing down the bias-variance decomposition of the quadratic mean square error.

$$E[(m_n(\underline{x}) - m(\underline{x}))^2] = \text{Var}(m_n(\underline{x})) + (E[m_n(\underline{x})] - m(\underline{x}))^2. \tag{2.26}$$

We reconsider the last expression in the early proof that is $\text{Var}(m_n(\underline{x})) = \frac{\gamma^2}{k(n)} (v(\underline{x}) + O(r))$. Then, we get $\text{Var}(m_n(\underline{x})) = \frac{\gamma^2}{k(n)} v(\underline{x})$ if r is small that is when $\frac{k(n)}{n}$ tends to zero. We achieve this by taking $k(n) = [n^Q]$ where $[\cdot]$ corresponds to the integer part of a real number. Therefore, it follows:

$$\text{Var}(m_n(\underline{x})) = O(n^{-Q}), \tag{2.27}$$

We have also the following expression of the bias from proof of point (i):

$$(E[m_n(\underline{x})] - m(\underline{x}))^2 = O(n^{-2\beta}). \tag{2.28}$$

Plugging equations (2.27) and (2.28) inside equation (2.26), we get $2\beta = Q$ or $Q = \frac{2p}{2p+d}$ and the proof is complete.

The result in theorem 2.1 provides knowledge of the bias and speed rate of the variance of the k -NN regression estimate in dependent case. Such knowledge allows for controlling finite sample

as well as asymptotic behavior of the estimator. Then, it ensures the use of such regression estimator to time series processes.

Knowledge of the asymptotic distribution of the k -NN regression estimate in dependent case is also important. We now study it.

2.2 Asymptotic normality

Still working with dependent processes, under the uniform mixing condition, we establish the asymptotic normality of k -NN regression estimate $m_n(\underline{x})$ in relation (2.1) as follows.

Theorem 2.2. *We suppose that assumptions 2.1-2.3 are verified. Then, we obtain the asymptotic normality of k -NN regression estimate $m_n(\underline{x})$ in relation (2.1) as follows:*

$$\sqrt{n^Q}(m_n(\underline{x}) - Em_n(\underline{x})) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (2.29)$$

where $0 \leq Q < 1$, $Q = \frac{2p}{2p+d}$, and $\sigma^2 = \gamma^2 v(\underline{x})$.

The result of theorem 2.2 is interesting because it permits the building of confidence interval which is fundamental in applications. Since confidence intervals can be used to compare the quality of point forecasts obtained from different methods, and enhances comparison of several methods (parametric and nonparametric methods), beyond point forecast. Therefore, from the asymptotic normality of the k -NN regression estimate $m_n(\underline{x})$ in Theorem 2.2, we build confidence interval whose expression is given in the following corollary.

Corollary 2.1. *Under the same assumptions as in theorem 2.2, a general form for the confidence interval around $m(\underline{x})$, for a given risk level $0 < \alpha < 1$, is:*

$$m(\underline{x}) \in [m_n(\underline{x}) - B - \frac{\hat{\sigma} z_{1-\frac{\alpha}{2}}}{\sqrt{k}}, m_n(\underline{x}) + B + \frac{\hat{\sigma} z_{1-\frac{\alpha}{2}}}{\sqrt{k}}] \quad (2.30)$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ quantile of the Student law, $\hat{\sigma}$ is an estimate for σ and B is such that:

1. B is negligible, if $\frac{k(n)}{n} \rightarrow 0$, as $n \rightarrow \infty$,
2. If not, $B = O(r^p)$, with $r = \left(\frac{k(n)}{(n-d)\hat{h}(\underline{x})c} \right)^{\frac{1}{d}}$ where $c = \frac{\pi^{d/2}}{\Gamma((d+2)/2)}$, and $\hat{h}(\underline{x})$ is an estimate for the density $h(\underline{x})$.

It usually happens that comparison of several point forecasts through the use of mean square prediction error is not sufficient. Corollary 2.1 allows us making comparison of forecast from k -NN regression with forecasts from other methods (parametric or nonparametric methods), beyond this standard mean square forecast prediction error. We now proceed to the proof of this corollary. We derive the proof of this corollary from the proof of theorem 2.2. So, We shall prove theorem 2.2.

Proof of Theorem 2.2. We assume that the variance $\sigma_n^2 = \text{var}[m_n(\underline{x})]$ exists and is non null, thus:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^{k(n)} \frac{w_i Y_i - Ew_i Y_i}{\sigma_n}. \quad (2.31)$$

To establish the asymptotic normality of $m_n(\underline{x})$, we distinguish two cases corresponding to the choice of the weighting functions.

i) The weights do not dependent on the process $(X_n)_n$, then equation (2.31) becomes:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^{k(n)} w_i Z_i, \quad (2.32)$$

where $Z_i = \frac{Y_i - EY_i}{\sigma_n}$. The asymptotic normality of equation (2.32) is obtained using theorem 2.2 in Peligrad and Utev (1997). To compute the variance, we use theorem 2.1. Then, we have $\sigma_n = \frac{\gamma^2}{k(n)} \text{var}(Y \mid \underline{X} = \underline{x})$. When we take $k(n) = [n^Q]$, therefore $\sigma^2 = \gamma^2 \text{var}(Y \mid \underline{X} = \underline{x})$ and the proof is complete.

ii) We assume that $w_i = \frac{w(\underline{x} - \underline{X}_{(i)})}{\sum_{i=1}^K w(\underline{x} - \underline{X}_{(i)})}$ where $w(\cdot)$ is a given function. In that latter case, the weights depend on the process $(X_n)_n$. In the following, we denote by $N(i)$ the order of the i^{th} neighbor. We rewrite the neighbor indices in an increasing order such that $M(1) = \min\{N(i), 1 \leq i \leq K\}$ and $M(k) = \min\{N(i) \notin \{M(j), \forall j < k\}, 1 \leq i \leq K\}$ for $2 \leq k \leq K$, and $K = k(n)$ is the number of neighbors. We introduce a real triangular sequence $\{\alpha_{Ki}, 1 \leq i \leq K \text{ and } \alpha_{Ki} \neq 0 \forall i\}$ such that

$$\sup_K \sum_{i=1}^K \alpha_{Ki}^2 < \infty \quad \text{and} \quad \max_{1 \leq i \leq K} |\alpha_{Ki}| \xrightarrow{n \rightarrow \infty} 0. \quad (2.33)$$

Now using the sequences $M(j), j = 1, \dots, K$ and $(\alpha_{Ki}), 1 \leq i \leq K$, we can rewrite expression

(2.31) as:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^K \alpha_{Ki} S_i, \quad (2.34)$$

with $S_i = \frac{w_{M(i)} X_{M(i)+1} - E w_{M(i)} X_{M(i)+1}}{\alpha_{Ki} \sigma_n}$. The sequence (S_i^2) is uniformly integrable and S_i is function only of $(X_j, j \leq M(i) + 1)$, thus if we denote \mathcal{F}_i , \mathcal{G}_i , \mathcal{F}_i^j and \mathcal{G}_i^j , the sigma algebras generated by $\{X_r\}_{r \leq i}$, $\{S_r\}_{r \leq i}$, $\{X_r\}_{r=i}^j$ and $\{S_r\}_{r=i}^j$ respectively, then $S_i \in \mathcal{F}_{M(i)+1}$, and $\mathcal{G}_i \subset \mathcal{F}_{M(i)+1}$. For a given integer ℓ , we have also $\mathcal{G}_{n+\ell}^\infty \subseteq \mathcal{F}_{n+M(\ell)+1}^\infty$ since $M(1) < M(1)+1 \leq M(2) < \dots \leq M(n+\ell) < M(n+\ell)+1 \leq M(n+\ell+1)$. Then:

$$\sup_{\ell} \sup_{A \in \mathcal{G}_1^\ell, B \in \mathcal{G}_{n+\ell}^\infty, P(A) \neq 0} |P(B | A) - P(B)| \leq \sup_{\ell} \sup_{A \in \mathcal{F}_1^{M(\ell)+1}, B \in \mathcal{F}_{n+M(\ell)+1}^\infty, P(A) \neq 0} |P(B | A) - P(B)|. \quad (2.35)$$

Under the α -mixing assumption on $(X_n)_n$, the right hand part of the expression (2.35) tends to zero as $n \rightarrow \infty$ and the left hand part of (2.35) converges to zero, hence the sequence $(S_i)_i$ is α -mixing. Moreover, for all i :

$$S_i \text{ is centered and } \text{var}\left(\sum_{i=1}^K \alpha_{Ki} S_i\right) = \text{var}\left(\frac{m_n(\underline{x})}{\sigma_n}\right) = 1. \quad (2.36)$$

Then, using expressions (2.33) - (2.36), and the theorem 2.2 in Peligrad and Utev (1997), we get:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} \rightarrow_{\mathcal{D}} \mathcal{N}(0, 1) \quad (2.37)$$

The variance of $m_n(\underline{x})$ is given by the relation (2.8). The proof of the theorem 2.2 is complete.

Next, we prove corollary 2.1.

Proof of corollary 2.1. From theorem 2.2, a confidence interval, for a given risk level α can be computed, and has the expression:

$$-z_{1-\frac{\alpha}{2}} \leq \frac{m_n(\underline{x}) - Em_n(\underline{x})}{\hat{\sigma}_n} \leq z_{1-\frac{\alpha}{2}} \quad (2.38)$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ quantile of Student law. Previously, we have established that the estimate $m_n(\underline{x})$ can be biased, thus the relationship (2.38) becomes:

$$m_n(\underline{x}) + B - \hat{\sigma}_n z_{1-\frac{\alpha}{2}} \leq m(\underline{x}) \leq m_n(\underline{x}) + B + \hat{\sigma}_n z_{1-\frac{\alpha}{2}} \quad (2.39)$$

When the bias is negligible, the corollary is established. If the bias is not negligible, we can bound it. The bound is obtained using expressions (2.15) and (2.40):

$$B = O\left(\left(\frac{k(n)}{(n-d)\hat{h}(\underline{x})c}\right)^{\frac{p}{d}}\right) \quad (2.40)$$

with $c = \frac{\pi^{d/2}}{\Gamma((d+2)/2)}$, $\hat{h}(\underline{x})$ being an estimate of the density $h(\underline{x})$. Introducing this bound in expression (2.39) completes the proof.

Some points can be mentioned, specially about the feasibility of the assumptions in theorem 2.1 and in theorem 2.2. We discuss these assumptions in the following remarks.

Remark 2.1. *As soon as the number of neighbors k is different from one, we remark that $\forall u$, $0 < w_i(u) < 1$, whatever the weighting function used (uniform or exponential function).*

Remark 2.2. *The main difference between k -NN method and kernel method lies on the information set that we use to estimate the function $m(\cdot)$ at a given point \underline{x} . In the latter case the information set is fix and in the former case, it is flexible with respect to the choice of the number of neighbors k . In this case, such a flexibility has an impact on the values of the weights. Indeed, when the number of neighbors k increases the weights $(w_i)_{i=1}^k$ decrease, then the product $(k \cdot w_i)_{i=1}^k$ turn around a constant γ which belongs to \mathbb{R} . For uniform weights, $w_i = \frac{1}{k}$ and $\gamma = 1$. This last property implies that the asymptotic variance of the estimate $m_n(\cdot)$ does not depend on the true density nor on the quantity $\int w^2(u)du$. This asymptotic property is not verified when we work with the kernel method.*

Remark 2.3. *The mixing conditions characterize different behaviors of dependent variables. Parametric processes like the bilinear models including ARMA models, the related GARCH processes and the Markov switching processes are known to be mixing, (Davydov 1973), (Guégan, 1983) and (Carrasco and Chen, 2002). Thus, in practice this condition is not too restrictive.*

Remark 2.4. *The condition in assumption 2.3 is verified in particular for the weights introduced in equation (2.2). The parameter γ introduced before entails the correlations between the vectors \underline{X}_n .*

3 Application on economic indicator modelling

We illustrate the theoretical result on asymptotic normality of the nearest neighbor regression method for computing confidence intervall. We focus on the modelling of the Euro Area monthly economic indicators. We then consider the following main economic indicators for such zone: the Industrial Production Index (IPI), the Industrial Production Index in Construction (IPIC)

where both indicators have 213 observations from January 1990 until September 2007, the Retail Sale Index (RI) with 153 observations from January 1995 to September 2007, the Confidence Indicator in Industry (ICI), the Confidence Index in Retail Trade (RCI) and the Consumer Confidence Indicator (CI) where the last four indicators have 275 observations from January 1985 until November 2007. We provide graphs of these six macroeconomic indicators in figure 1 in the appendix.

We couple our study with the classical linear parametric model frequently considered as benchmark in the modelling of economic indicators. Then, we consider the linear ARMA modelling for the six economic indicators and if necessary use the nonlinear GARCH process for modelling the conditional variance. For each indicator, we build the model on the sample without the last two year observations. For stationary condition, we take the first order difference for the six time series. Based on the standard serial correlation order selection, we retain the following models for the six indicators: an ARIMA(3,1,0) for IPI, an ARIMA(2,1,0)-ARCH(1) for IPIC, an ARIMA(3,1,0) for ICI, an ARIMA(2,1,0)-GARCH(1,1) for RCI, an ARIMA(2,1,0) for RI and an ARIMA(3,1,0) for CI. In the case of the k -NN regression method; we determine the two parameters, the embedding dimension d and the number of neighbors k , based on the minimization of the root mean square error using the same transformation on the time series as in AR modelling. We then obtain $kNN(4, 2)$ for IPI, $kNN(6, 2)$ for IPIC, $kNN(10, 3)$ for ICI, $kNN(10, 4)$ for RCI, $kNN(6, 5)$ for RI and $kNN(9, 2)$ for CI where in the notation $kNN(·, ·)$ the first coordinate corresponds to the embedding dimension d and the second to the number of neighbors k . Moreover, we have considered the exponential weighting function since it reflects the local behavior of nearest neighbors method giving more weight to closest neighbors.

We now provide confidence intervals for two year forecasts computed from the previous two regression methods. We make use of the theoretical results on confidence intervals, developed in previous section, for the forecasts based on kNN regression method. We mention that the required assumptions for the application of these results are verified for our data sets: the stationary condition, mixing condition from remark 2.3 and the condition on the weighting function. Thus, we can use these results to build the confidence intervals for k -NN regression method. The confidence intervals at 95% for forecasts for both k -NN regression and ARIMA modelling are

given in figure 2 and figure 3 in the appendix.

Concerning the pointwise forecasts from the two methods (ARMA modelling and k -NN method), we observe that the ARMA modelling provides the classical mean forecast as the horizon increases (curve in star). The k -NN forecasts (curve in plus sign) follow the true trajectory (dotted curve) except for the ICI indicator. Next, concerning the confidence intervals from both methods, in general the true trajectory stays always inside the k -NN confidence intervals (curves in circle). This is not always the case for the confidence intervals from the linear ARMA model (curves in triangle). We can also notice that the bandwith of the k -NN confidence interval seems smaller than the bandwith of the ARMA confidence interval in particular for smaller forecast horizons h (h corresponding to one year). Thus, we obtain comparable forecast for both k -NN method and ARMA-GARCH modelling for point foreacast as well as confidence forecast.

4 conclusion

Two main results have been obtained when working beyond independent processes. First, we have characterized the bias, the variance, the quadratic mean square error of the k -NN regression estimate of the conditional mean. For strongly mixing processes with suitable conditions, the bias and the quadratic mean square error of the k -NN regression estimate are of order $O(n^{-\beta})$ and of order $O(n^{-Q})$, respectively. Second, we have established asymptotic normality of the k -NN regression estimate of the conditional mean and derived confidence interval for such regression estimate. The asymptotic normality of the same k -NN regression estimate has been provided where the centered k -NN regression estimate scaled by the corresponding asymptotic variance has asymptotic standard normal distribution with speed rate of convergence $\sqrt{n^Q}$.

References

- [1] Bosq, D. (1998), “Nonparametric Statistics for Stochastic Processes: Estimation and Prediction”, Lecture Notes in Statistics, Springer-Verlag, 2nd eds, Berlin.
- [2] Bradley, R.C. (1986), “Basic properties of strong mixing conditions”, In Dependence in Probability and Statistics, E. Eberlein and M.S. Taqqu (eds.), 165-192, Birkhauser, Boston.

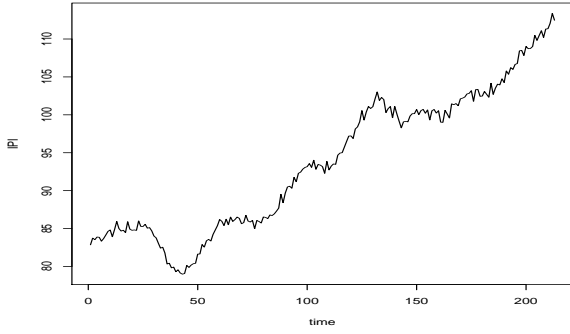
- [3] Carrasco, M. and X. Chen (2002), “Mixing and moment properties of various GARCH and Stochastic Volatility models”, *Econometric Theory* 18, 17-39.
- [4] Collomb, G. (1984), “Nonparametric time series analysis and prediction: Uniform almost sure convergence of the window and k-NN autoregression estimates”, *Mathematic Operational Statistics, Series Statistics*.
- [5] Davydov Y.A. (1973), “ Mixing conditions for Markov chains”, *Theory Probab. Appl.* 18, 312-328.
- [6] Doukhan, P. (1994), “Mixing: properties and examples”, *Lecture Notes in Statistics* 85, Springer Verlag.
- [7] Devroye, L.P. (1982), “Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates”, *Z. Wahrscheinlichkeitstheorie Verw Gebiete* 61, 467-481.
- [8] Girosi, F. and G. Anzellotti (1992), “Rates of convergence of approximation by translates”, A.I. Memo 1288, Artificial Intelligence Laboratory, MIT.
- [9] Guégan, D. (1983), “Une condition d’ergodicité pour des modèles bilinéaires à temps discret”, *CRAS Série 1*, 297-301.
- [10] Ibragimov, I.A. (1962), “Some limit theorems for stationary processes” *Theory of Probability and Application* 7, 349-382.

Lin, Z.Y. and C.R. Lu (1996), “Limit theory for mixing dependent random variables”, Kluwer Academic Publishers.
- [11] Mack, Y.P. (1981), “Local properties of k -NN regression estimates”, *SIAM Journal on Algebraic and Discrete Methods* 2, 311-323.
- [12] Mokkadem, A. (1990), “Propriétés de mélange des modèles autoregressifs polynomiaux”, *Annales de l’Institut Henri Poincare* 26, 219-260
- [13] Peligrad, M. (1986), “Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey)”, In *Dependence in Probability and Statistics*, E. Eberlein and M.S. Taqqu (eds.), 193-223, Birkhauser, Boston.

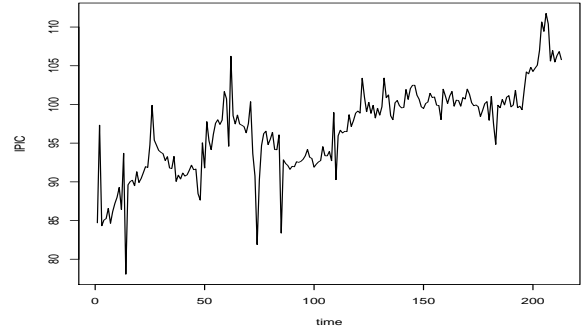
- [14] Peligrad, M. and S.A. Utev (1997), “Central limit theorem for Linear Processes”, The Annals of Probability 25, 443-456.
- [15] Rosenblatt, M. (1956), “A central limit theorem and a strong mixing condition”, Proc. Natl. Acad. Sci. USA 42, 43-47.
- [16] Royall, R. M. (1966), “A class of nonparameteric estimates of a smooth regression function”, Ph.D. dissertation, Stanford University, Stanford.
- [17] Stone, C. (1977), “Consistent non parametric regression”, Annals of Statistics 5, 595-645.
- [18] Yakowitz, S. (1987), “Nearest neighbors method for time series analysis”, Journal of Time Series Analysis 8, 235-247.

5 Appendix

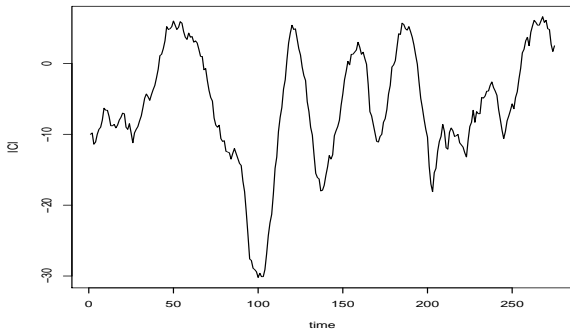
Figure 1: Six macroeconomic indicators of the Euro Area.



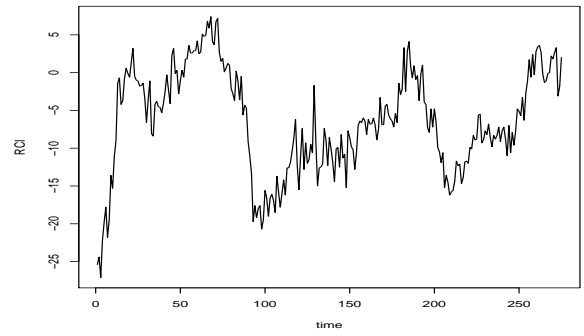
(a) Monthly IPI from 01-1990 to 09-2007



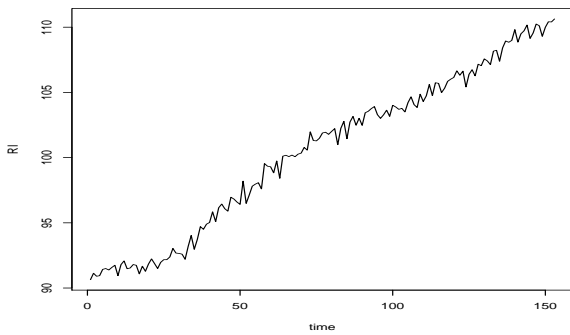
(b) Monthly IPIC from 01-1990 to 09-2007



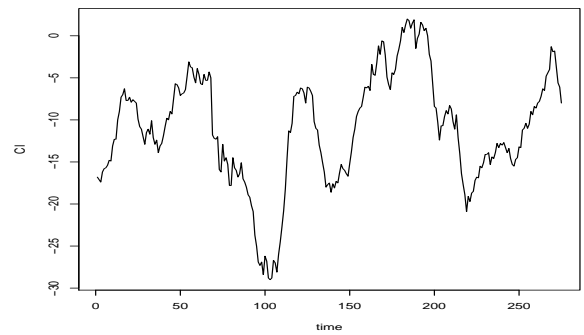
(c) Monthly ICI from 01-1985 to 11-2007



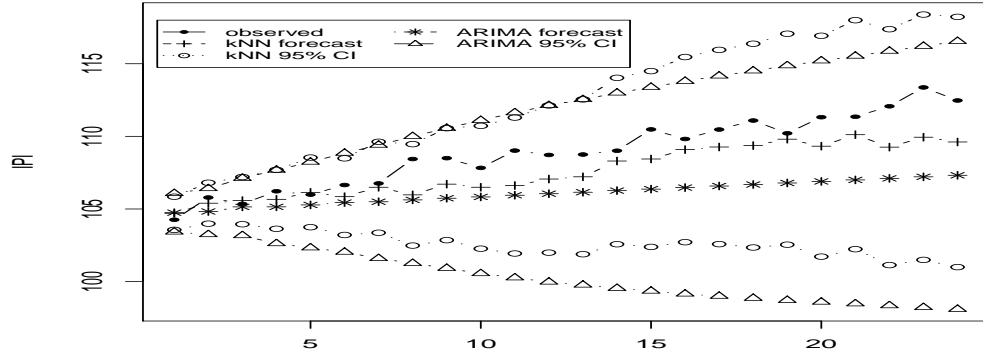
(d) Monthly RCI from 01-1985 to 11-2007



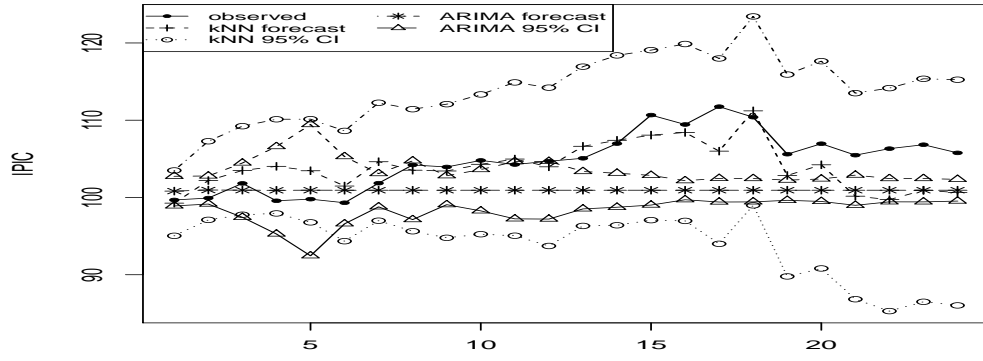
(e) Monthly RI from 01-1995 to 09-2007



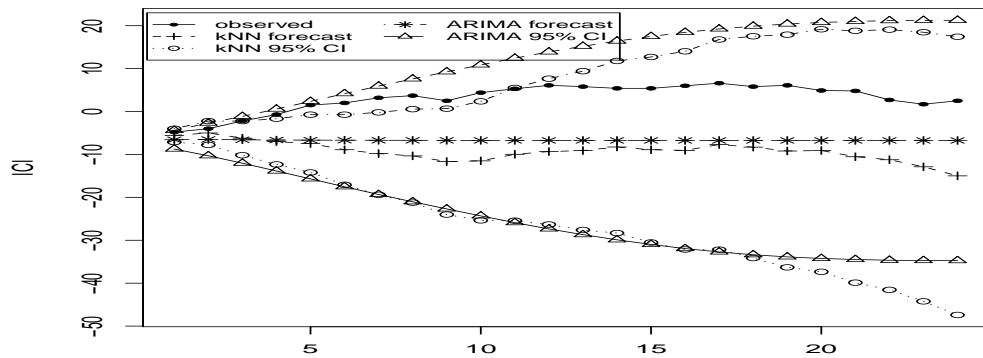
(f) Monthly CI from 01-1985 to 11-2007



(g) 95% confidence interval of IPI forecast by k -NN and ARMA-GARCH from 10-2005 to 09-2007.

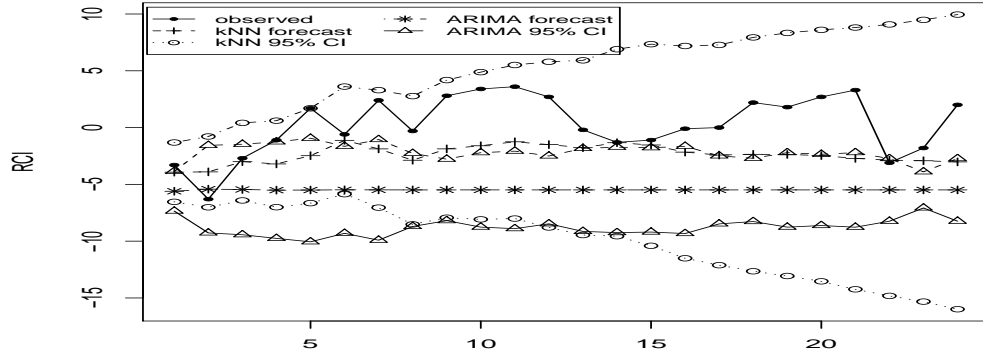


(h) 95% confidence interval of IPIC forecast by k -NN and ARMA-GARCH from 10-2005 to 09-2007.

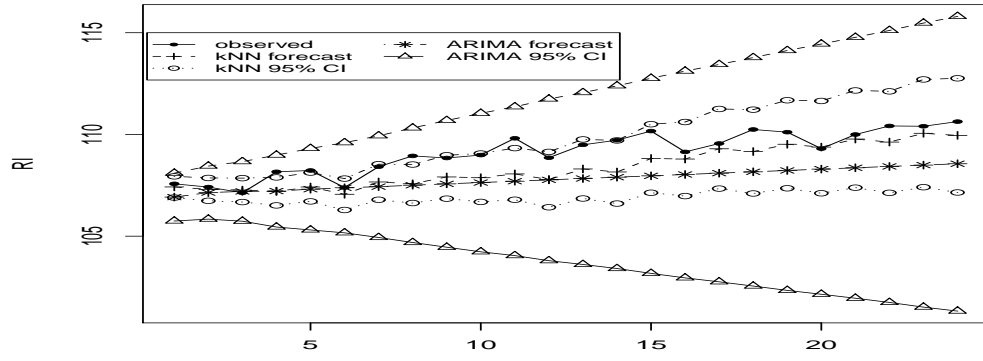


(i) 95% confidence interval of ICI forecast by k -NN and ARMA-GARCH from 12-2005 to 11-2007.

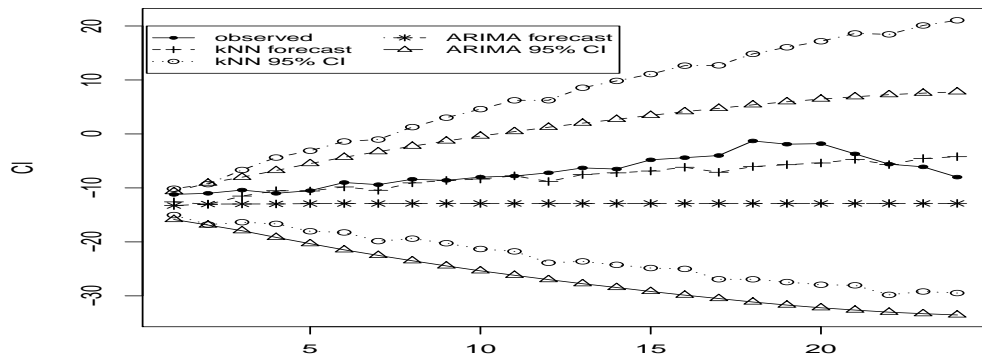
Figure 2: IPI, IPIC and ICI observed values (dashed) and forecasted values by k -NN (plus sign) and by ARMA-GARCH (star) with their 95% Confidence intervals (circle) and (triangle) respectively.



(a) 95% confidence interval of RCI forecast by k -NN and ARMA-GARCH from 12-2005 to 11-2007.



(b) 95% confidence interval of RI forecast by k -NN and ARMA-GARCH from 10-2005 to 09-2007.



(c) 95% confidence interval of CI forecast by k -NN and ARMA-GARCH from 12-2005 to 11-2007.

Figure 3: RCI, RI and CI observed values (dashed) and forecasted values by k -NN (plus sign) and by ARMA-GARCH (star) with their 95% confidence intervals (circle) and (triangle) respectively.